# Food Calorie Estimation and Auto Bill Generation for Grocery Products using YOLO Object Detection

*Ms. Y. Vineela Sravya*
*Assistant Professor,*
*Department Of Computer Science and Engineering*
*Vignan's Institute of Engineering for Women*
*Vishakhapatnam, Andhra Pradesh, India*

*M. Keerthi*
*Department Of Computer Science and Engineering*
*Vignan's Institute of Engineering for Women*
*Vishakhapatnam, Andhra Pradesh, India*

*M. Kasturi*
*Department Of Computer Science and Engineering*
*Vignan's Institute of Engineering for Women*
*Vishakhapatnam, Andhra Pradesh, India*

*R. Lochana*
*Department Of Computer Science and Engineering*
*Vignan's Institute of Engineering for Women*
*Vishakhapatnam, Andhra Pradesh, India*

*A. Anusha*
*Department Of Computer Science and Engineering*
*Vignan's Institute of Engineering for Women*
*Vishakhapatnam, Andhra Pradesh, India*

**ABSTRACT** – In today's world image processing, is a key for object detection as it has became a challenging task in various application domains. Earlier bar-code technology is used for bill generation which takes a lot of time and can only be read one item individually, which is the major disadvantage and a time-consuming process. In order to overcome the existing technology and generate bill within less time, we proposed an automatic bill generation system which displays the price and calorie of the chosen grocery products during bill generation and estimates calorie count and price for the food item at one glance. To attain this, a dataset is made for product price and food calorie which is used at the time of billing. By utilizing YOLO algorithm it's easy to detect and recognize the objects much faster when compared with CNN and other algorithms. As CNN cannot encode the position and orientation of the object i.e. it cannot identify an image which is rotated, sized differently or viewed in different illumination.

***Keywords***: CNN, YOLO, image processing, object detection.

## I. INTRODUCTION

In the era of digital world, image processing plays an important role as images contain a lot of information. By using images we aim to develop an automatic billing system that can be used in supermarkets which helps the customers to check out with their grocery products in less time rather than waiting in lines for hours in the sophisticated environment. With the emergence of new technologies, like one-click bill generator and wireless networks, which

make the shopping, process faster, transparent and efficient. Here our billing system is equipped with a camera for product identification with the shop's server. Moreover, it also has an LCD display that informs the shop retailer and customers about total price and total calorie. As soon as the object is detected, it identifies the product and updates the bill along with calorie count. The billing system will change the way people shop as radically as ATM's changed banking. The proposed system is easy to use and does not need any special training to operate the system. Here the automatic billing system is connected to a camera that gives input an image to generate bill which makes shopping a breeze and has other positive spin-offs such as freeing staff from repetitive checkout scanning.

### A. Deep learning

Deep learning is used for data mining i.e., in the field of data science. It consists of networks capable of learning unsupervised from data that is unstructured or unlabeled. Deep learning is an artificial intelligence function. It is similar to the working of the human brain in processing data and creating patterns to make decisions. Deep learning tells about the future where machine learning tells about the present.

The output of deep learning can be anything a number, an element, sound, text, etc. In deep learning training takes more times when compared with machine learning. Deep learning performs automatic feature extraction without the need for human intervention.

### B. YOLO (You Only Look Once)

It is a mixture of Convolutional Neural Networks (CNN) and deep learning for object detection; it is different from other algorithm because, as the name indicates it only needs to "see" each image once. This makes YOLO to be the fastest object detection algorithm (naturally sacrificing some accuracy). Here human doesn't need to select or crop the required region from the image. YOLO detects multiple objects in an image only at a single run of the algorithm with class probabilities and bounding boxes using single neural network. In real time it can process 45 frames per second.

### C. CNN (Convolutional Neural Networks)

The CNN is one of the algorithms used for image recognition. It is a model for building a hierarchical network which, finally gives out a fully-connected layer where all the neurons are connected to each other and the output is processed.

### Layers in CNN

**Input Layer:** The main task of input layer is to initialize the input image data to make all the dimensions of the input data zero-centred. Then, it normalizes the dimensions of all the input data within [0, 1] to accelerate the converging speed.

**Convolutional (CONV) Layer**: Is the core of CNNs which is used for sliding on the image. The values of each pixel within the slide will be multiplied and added as the convolutional results.

**Active Layer:** To make the results of the convolution layer nonlinear, active Layer is about solving the vanishing gradient problem caused by the under fitting. In our yolo algorithm Leaky ReLU function is used which calculates in a simpler and more efficient way, and contains no dead area as it is faster than sigmoid function.

**Pooling Layer:** It is used to reduce the dimension of the results from convolution layer, which is in between two convolution layers. It's typical ways are max pooling and average pooling.

**Fully Connected Layer**: It is the last layer in CNN which is gives the output with speeds the calculations.

## II. LITERATURE SURVEY

In our literature survey [1] Joseph Redmon and others who proposed Unified Real-Time Object Detection which is also known as YOLO (You Only Look Once) have worked on object detection using a regression algorithm in order to get high accuracy and good predictions. [2] "Understanding of object detection based on CNN Family and YOLO" by Juan Du, explained about the object detection algorithms like CNN, R-CNN and compared their efficiency and introduced YOLO algorithm to increase the efficiency. [3] "Object Localization with Structured Output Regression", by Matthew B. Blaschko has used the bounding box method for object localization to overcome the drawbacks of the sliding window method.

## III. BUILDING A DATASET

### A. FOR FOOD IMAGES:

Here we considered 1000 images for each food item for the evaluation of food item. In order to recognize the food item, each food item region is needed to be identified and stored in the dataset. The user takes a photo of a of there meal or a snack region which is involving a food item via the touch panel display of the smart phone by in- putting the name of the food item. The food item name is shown on the bill from the food database.

Figure 1. Example dataset

### B. FOR PRICE & CALORIE

In this, a dataset is created to display price and calorie for the detected product. If the detected product is a food item it estimates calorie count and also price. If the detected product is not a food item then it only estimates the price.

Table 1: Dataset for price and calorie estimation

| Item | Price | Calorie |
|---|---|---|
| Burger | 100 | 295 |
| Samosa | 20 | 262 |
| Doughnut | 125 | 452 |
| Noodles | 60 | 138 |
| Pani Puri | 50 | 329 |
| Fried Rice | 80 | 163 |
| Soap | 85 | - |
| Shampoo | 200 | - |
| Pen | 15 | - |
| Book | 40 | - |

## IV. ALGORITHM USED IN PROSOPED SYSTEM

In YOLO algorithm an image is passed as an input for detection which is applied on the image. For example, consider the burger image below which is divided into 3x3 matrixes normally we can divide the image into any number of grids, based on the image intricacy. After the image is divided, each grid is classified and localized. Then a confidence score of each grid is calculated which shows the accuracy value. If a grid doesn't contain an object then the bounding box value of the grids are going to be zero or if an object is found within the grid then the bounding box value is going to be 1.

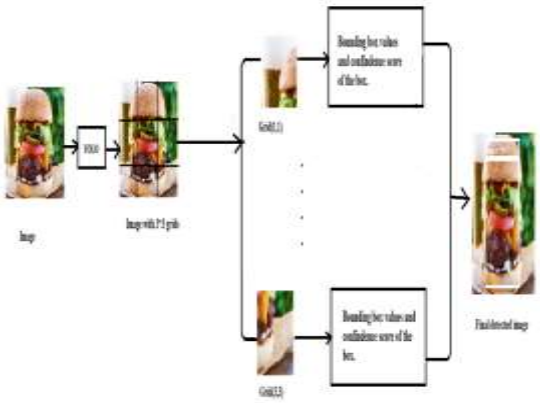

Figure 2. Yolo working model

### A. Bounding box predictions:

In yolo bounding box plays a vital role where it divides an image into M X M grids. Then for every grid classification & object localization is applied and each grid is labelled. Later algorithm checks for every grid separately and labels the object with its bounding boxes. The girds without object are marked as zero.

Figure 3. Example image with 3x3 grids

## B. ANCHOR BOX

With the help of bounding boxes yolo can only detect single object, from a grid. So, in order to detect more than one object the yolo algorithm uses anchor box.

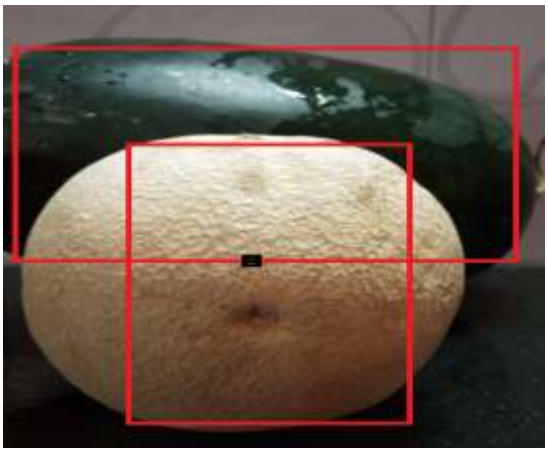

Figure 4. An example image for anchor boxes

Consider the above figure, both watermelon and the muskmelon's midpoint fall on the same grid. Those rectangles are the anchor boxes for the above fruits. A number of anchor boxes are generated from a single image when the algorithm detects multiple objects. In above image we have 2 anchor boxes.

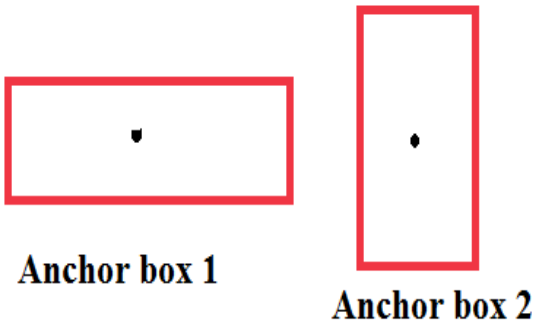

Figure 5. Anchor boxes

In above image we have 2 anchor boxes. The anchor box which is vertically drawn is for the watermelon and the horizontally drawn box is for muskmelon.

There will be 16 values as there are to anchor boxes as an overlapping object is detected.

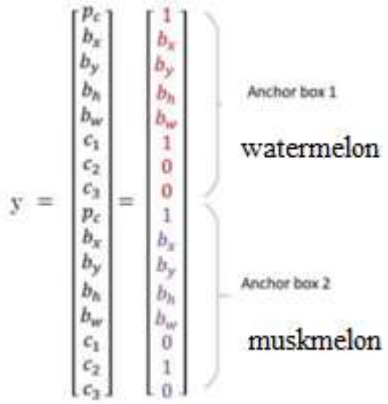


Figure 6. Prediction values for both anchor boxes

**pc** it represents whether the object is present or not in a particular anchor box.

**bx, by, bh, bw** are the bounding box values for both the anchor boxes.

The class value for the 1st anchor box i.e., (c1, c2, c3) is (1, 0, 0) because the detected object is muskmelon.

In the 2nd anchor box, the detected object is a watermelon so the class value will be (0, 1, 0).

Here the matrix form of Y is 3 X 3 X 16 we can also write 3 X 3 X 2 X 8 as there are two anchor boxes.

## V. EXPERIMENTAL RESULT

The result obtained from the above system is that after capturing the image of the product or the food item, by using YOLO object detection method. It will generate bill which is displayed on the LCD screen using the database where we have stored each and every detail of the product and the food item such as its name, price and also the cost of grocery products and price and calorie for the food item.

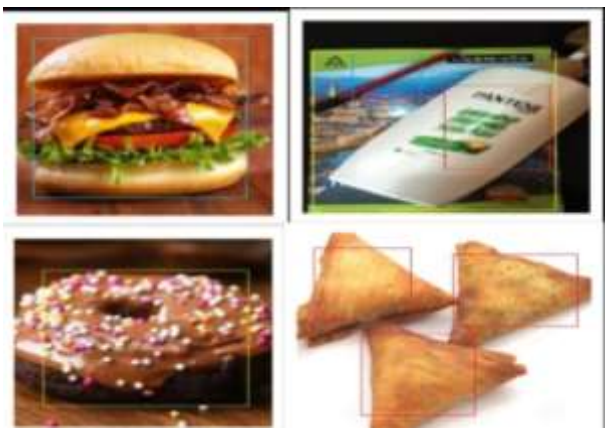

Figure 7. The above rectangle boxes are bounding boxes.

Figure 8. Output which is displayed on the LCD screen

## VI. CONCLUSION

In conclusion we would like to make our proposed system into a commercially viable product as an excellent way to help our customers by reducing the time spent in shopping by displaying the list of products, their cost and calorie intake at the time of billing. We also found that YOLO's accuracy was faster and efficient than traditional methods.

As future work, we plan to develop an application which suggest user the amount of calories to be consumed in a day and also send a message which shows the total amount and total calorie of their purchased products.

## VII. REFERENCES

1. "You Only Look Once: Unified Real-Time Object Detection" by Joseph Radmon, Sathosh Divvala and Ros Girsick in IEEE Conference on CVPR 2016.
2. YOLO Juan Du1, wrote"Understanding of Object Detection Based on CNN", NRDCOH Qingdao 266071, China.
3. "Learning to Localize Objects with Structured Output Regression" by Matthew B. Blaschko Christ H. Lampert, which is published in Computer Vision 2008.
4. "SSD: Single Shot MultiBox Detector", by Wei liu and others published in CV 2016.
5. "Unifying Landmark Localization with End to End Object Detection" by Lichao, Yi Yang, Yafeng D, Yinan Yu, published in CVPR.
6. "scalable object detection using DNN" by Dumitru Erhan, Christian Szegedy, Alexander Toshev, In the IEEE Conference on CVPR in 2014.
7. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", published in ANIPS in 2015.
8. YOLO9000 by Joseph Redmon, Ali Farhadi, In IEEE Conference on (CVPR) 2017.
9. "R-FCN: Object Detection via Region-based Fully Convolutional Networks", by Jifeng Dai, Yi Li, Kaiming He and Jian Sun published in ANIPS 2016.
10. "Very Deep Convolutional Networks for Large-Scale Image Recognition", by Karen simonyan and Andrew zisserman published in CV&RP.